Identifying Multimodal Errors

Through Explanations

Dr. Leilani H. Gilpin
Research Scientist, Sony AI
December 8, 2020

Brief Introduction

- Joined Sony AI in September 2020.
- Education
  - B.S. in Computer Science, B.S. in Mathematics at UCSD (2011).
  - M.S. in Computational Math from Stanford University (2013).
  - PhD in EECS from MIT (2020).
- Career Path
  - Worked at Xerox PARC from 2013-2015
  - Started Sony AI immediately after my PhD.
- Current Position: Research Scientist
- Details on Job
  - Adding explainability to AI agents.
  - Explanations for diagnosis and debugging.

Local Sanity Checks

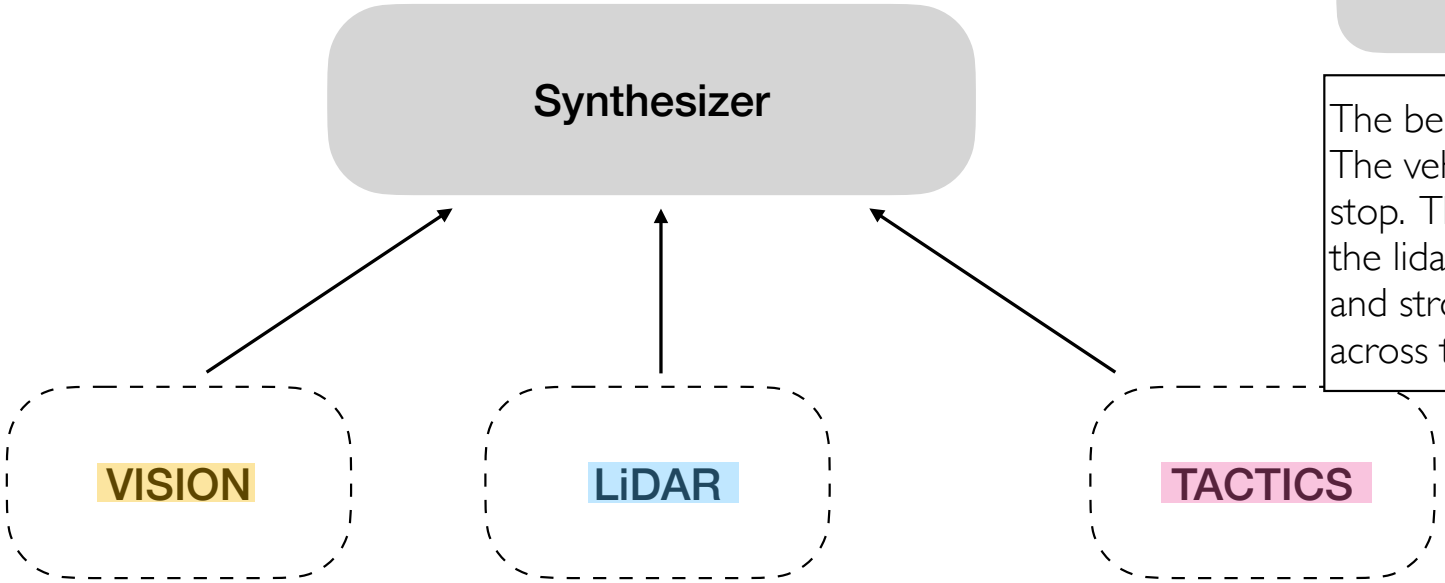
Synthesizer to reconcile inconsistencies between parts.

A Deadly Crash

Solution: Internal Communication

Anomaly Detection through Explanations

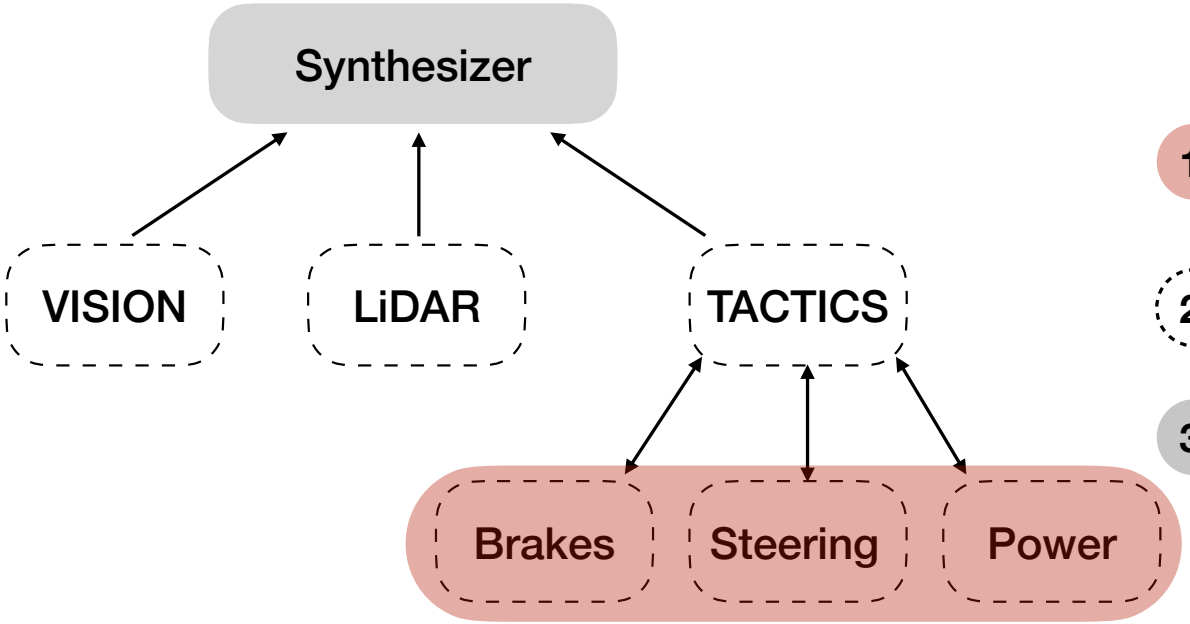
Synthesizer to reconcile inconsistencies between monitor outputs.



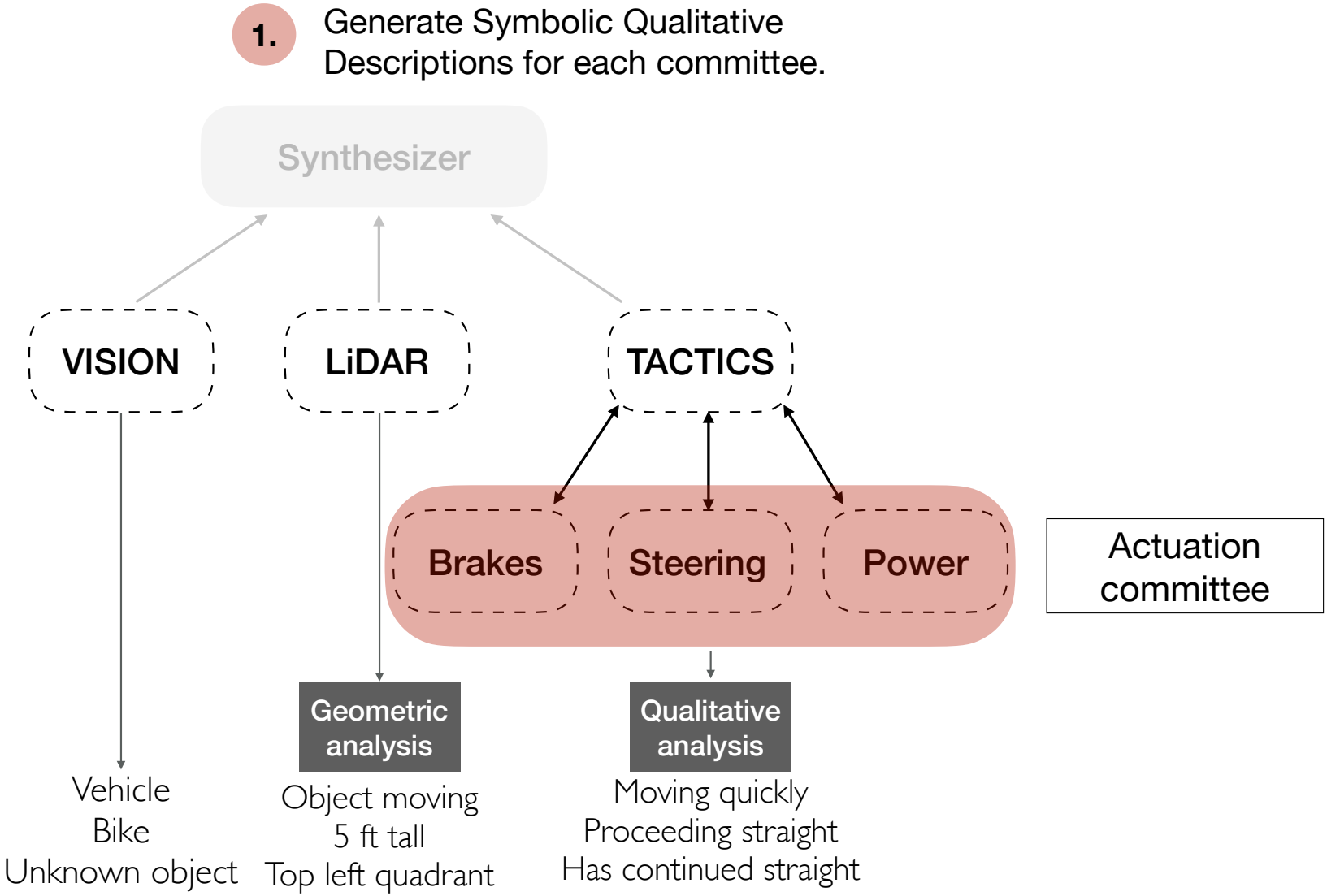
The best option is to veer and slow down. The vehicle is traveling too fast to suddenly stop. The vision system is inconsistent, but the lidar system has provided a reasonable and strong claim to avoid the object moving across the street.

Anomaly Detection through Explanations

Reasoning in Three Steps

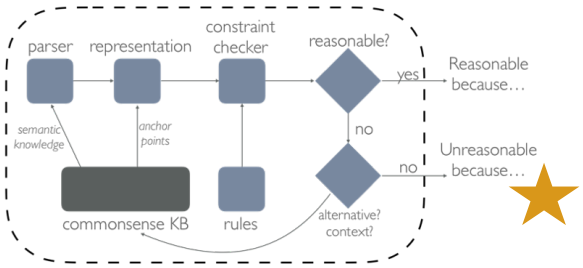


1. Generate Symbolic Qualitative Descriptions for each committee.
2. Input qualitative descriptions into local "reasonableness" monitors.
3. Use a synthesizer to reconcile inconsistencies between monitors.


1.
Generate Symbolic Qualitative Descriptions for each committee.
Synthesizer
VISION
LiDAR
TACTICS
Brakes
Steering
Power
Actuation committee
Geometric analysis
Qualitative analysis
Vehicle
Bike
Unknown object
Object moving
5 ft tall
Top left quadrant
Moving quickly
Proceeding straight
Has continued straight

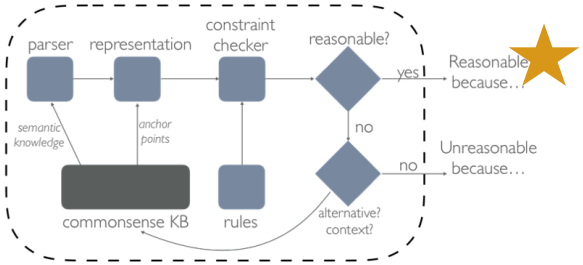
2. Input qualitative descriptions into local "reasonableness" monitors.

Vehicle
Bike
Unknown object



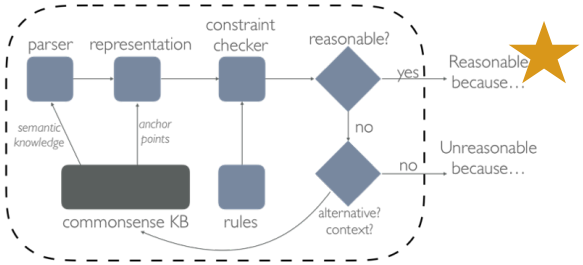
This vision perception is unreasonable. There is no commonsense data supporting the similarity between a vehicle, bike and unknown object except that they can be located at the same location. This component's output should be discounted.

Object moving
5 ft tall
Top left quadrant



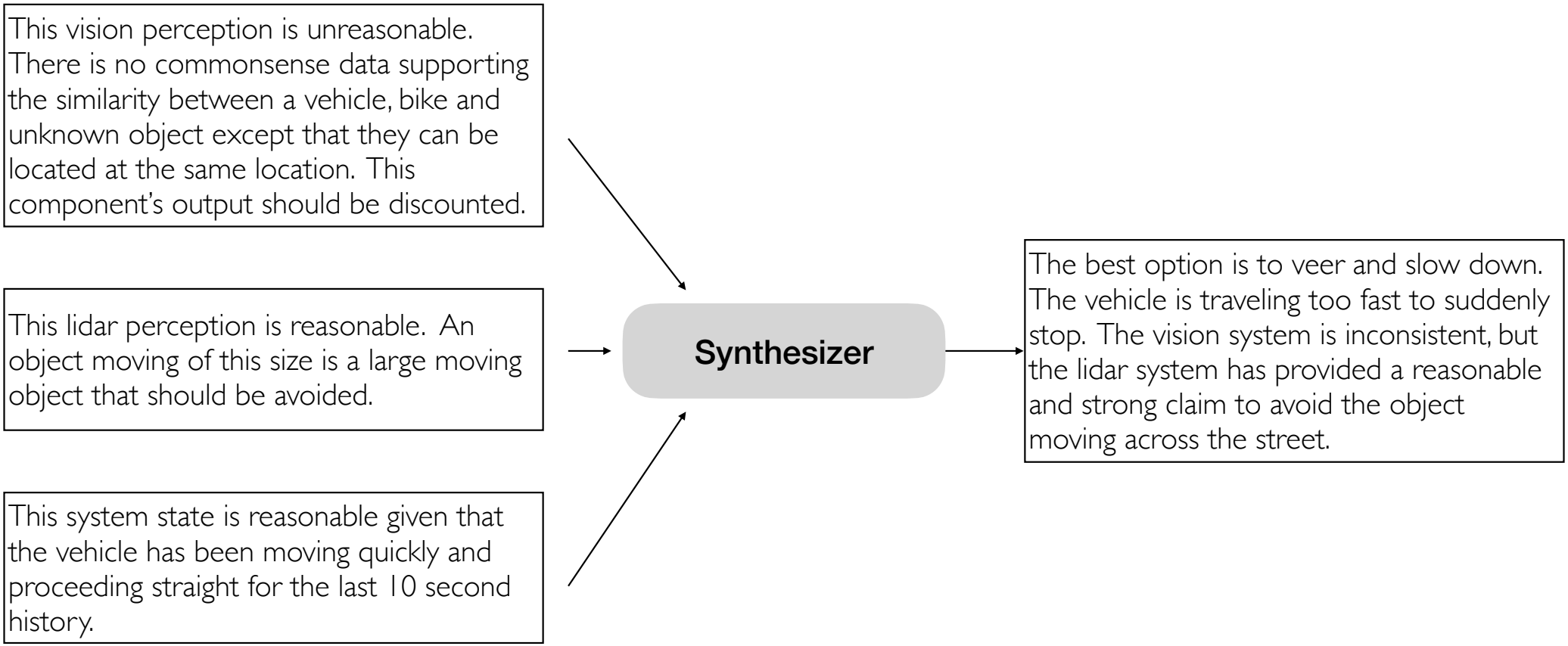
This lidar perception is reasonable. An object moving of this size is a large moving object that should be avoided.

Moving quickly
Proceeding straight
Has continued straight



This system state is reasonable given that the vehicle has been moving quickly and proceeding straight for the last 10 second history.

3. Use a synthesizer to reconcile inconsistencies between monitors.



3. Use a synthesizer to reconcile inconsistencies between monitors.

Symbolic reasons

```
(monitor, judgement, unreasonable)
(input, isType, labels)
(all_labels, inconsistent, negRel)
(isA, hasProperty, negRel)
…
(all_labels, notProperty, nearMiss)
(all_labels, locatedAt, consistent)
(monitor, recommend, discount)
```

```
(monitor, judgement, reasonable)
(input_data, isType, sensor)
…
(input_data[4], hasSize, large)
(input_data[4], IsA, large_object)
(input_data[4], moving, True)
(input_data[4], hasProperty, avoid)
```

```
(monitor, judgement, reasonable)
(input, isType, history)
(input_data, moving, True)
(input_data, direction, forward)
(input_data, speed, fast)
(input_data, consistent, True)
(monitor, recommend, proceed)
```

Synthesizer

The best option is to veer and slow down. The vehicle is traveling too fast to suddenly stop. The vision system is inconsistent, but the lidar system has provided a reasonable and strong claim to avoid the object moving across the street.

3. Use a synthesizer to reconcile inconsistencies between monitors.

Synthesizer + Priority Hierarchy → Abstract Goals

- Explanation synthesizer to deal with *inconsistencies.*
  - Argument tree.
  - Queried for support or counterfactuals.

1. Passenger Safety
2. Passenger Perceived Safety
3. Passenger Comfort
4. Efficiency (e.g. Route efficiency)

→ A passenger is safe if:

- The vehicle proceeds at the same speed and direction.
- The vehicle avoids threatening objects.

3. Use a synthesizer to reconcile inconsistencies between monitors.

```
(monitor, judgement, unreasonable)
(input, isType, labels)
(all_labels, inconsistent, negRel)
(isA, hasProperty, negRel)
…
(all_labels, notProperty, nearMiss)
(all_labels, locatedAt, consistent)
(monitor, recommend, discount)
```

```
(monitor, judgement, reasonable)
(input, isType, sensor)
…
(input_data[4], hasSize, large)
(input_data[4], IsA, large_object)
(input_data[4], moving, True)
(input_data[4], hasProperty, avoid)
…
(monitor, recommend, avoid)
```

!

```
(monitor, judgement, reasonable)
(input, isType, history)
(input_data, moving, True)
(input_data, direction, forward)
(input_data, speed, fast)
(input_data, consistent, True)
(monitor, recommend, proceed)
```

Abstract Goal Tree

```
'passenger is safe',
AND(
  'safe transitions',
  NOT('threatening objects')
```

!

The best option is to veer and slow down. The vehicle is traveling too fast to suddenly stop. The vision system is inconsistent, but the lidar system has provided a reasonable and strong claim to avoid the object moving across the street.

Evaluation in Simulation